

IPW

PTO/SB/21 (08-00)

Approved for use through 10/31/02. OMB 0651-0031

Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays valid OMB control number.

TRANSMITTAL FORM <i>(to be used for all correspondence after initial filing)</i>	Application Number	10/676,387	
	Filing Date	September 30, 2003	
	First Named Inventor	Gabriel Monaton, et al.	
	Group Art Unit	2123	
	Examiner Name	NYA	
Total Number of Pages in This Submission	32	Attorney Docket Number	15437-0635

ENCLOSURES (check all that apply)		
<input type="checkbox"/> Fee Transmittal Form	<input type="checkbox"/> Assignment Papers (for an Application)	<input type="checkbox"/> After Allowance Communication to Group
<input type="checkbox"/> Fee Attached	<input type="checkbox"/> Drawing(s)	<input type="checkbox"/> Appeal Communication to Board of Appeals and Interferences
<input type="checkbox"/> Amendment / Response	<input type="checkbox"/> Licensing-related Papers	<input type="checkbox"/> Appeal Communication to Group (Appeal Notice, Brief, Reply Brief)
<input type="checkbox"/> After Final	<input type="checkbox"/> Petition	<input type="checkbox"/> Proprietary Information
<input type="checkbox"/> Affidavits/declaration(s)	<input type="checkbox"/> Petition To Convert To a Provisional Application	<input type="checkbox"/> Status Letter
<input type="checkbox"/> Extension of Time Request	<input type="checkbox"/> Power of Attorney, Revocation Change of Correspondence Address	<input type="checkbox"/> Other Enclosure(s) (please identify below):
<input type="checkbox"/> Express Abandonment Request	<input type="checkbox"/> Terminal Disclaimer	
<input type="checkbox"/> Information Disclosure Statement	<input type="checkbox"/> Request for Refund	
<input checked="" type="checkbox"/> Certified Copy of Priority Document(s)	<input type="checkbox"/> CD, number of CD(s) _____	
from France dtd 13 Juin 2003	Remarks	
<input type="checkbox"/> Response to Missing Parts/ Incomplete Application		
<input type="checkbox"/> Response to Missing Parts under 37 CFR 1.52 or 1.53		

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT	
Firm or Individual name	Hickman Palermo Truong & Becker LLP Craig G. Holmes
Signature	
Date	June 3, 2004

CERTIFICATE OF MAILING			
I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class: mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on this date: 6/3/04			
Type or printed name	Annette Jacobs		
Signature		Date	June 3, 2004

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

THIS PAGE BLANK (USPTO)

02/12/07
netter (1)

BREVET D'INVENTION

CERTIFICAT D'UTILITÉ - CERTIFICAT D'ADDITION**COPIE OFFICIELLE**

Le Directeur général de l'Institut national de la propriété industrielle certifie que le document ci-annexé est la copie certifiée conforme d'une demande de titre de propriété industrielle déposée à l'Institut.

Fait à Paris, le **13 JUIN 2003**

Pour le Directeur général de l'Institut
national de la propriété industrielle
Le Chef du Département des brevets

Martine PLANCHE

**INSTITUT
NATIONAL DE
LA PROPRIÉTÉ
INDUSTRIELLE**

SIEGE
26 bis, rue de Saint Petersburg
75800 PARIS cedex 08
Téléphone : 33 (0)1 53 04 53 04
Télécopie : 33 (0)1 53 04 45 23
www.inpi.fr

THIS PAGE BLANK (USPTO)



26 bis, rue de Saint Pétersbourg
75800 Paris Cedex 08
Téléphone : 01 53 04 53 04 Télécopie : 01 42 94 86 54

BREVET D'INVENTION**CERTIFICAT D'UTILITÉ**

Code de la propriété intellectuelle - Livre VI



N° 11354*01

REQUÊTE EN DÉLIVRANCE 1/2**Important !** Remplir impérativement la 2ème page.

Cet imprimé est à remplir lisiblement à l'encre noire

09 540 W / 190600

Réserve à l'INPI REMISE DES PIÈCES DATE 30 SEPT 2002 LIEU 75 INPI PARIS N° D'ENREGISTREMENT 0212075 NATIONAL ATTRIBUÉ PAR L'INPI DATE DE DÉPÔT ATTRIBUÉE PAR L'INPI 30 SEP. 2002		1 NOM ET ADRESSE DU DEMANDEUR OU DU MANDATAIRE À QUI LA CORRESPONDANCE DOIT ÊTRE ADRESSÉE CABINET NETTER 36 avenue Hoche 75008 PARIS	
Vos références pour ce dossier (facultatif) SUN 43 (120749)			
Confirmation d'un dépôt par télécopie <input type="checkbox"/> N° attribué par l'INPI à la télécopie			
2 NATURE DE LA DEMANDE		Cochez l'une des 4 cases suivantes	
Demande de brevet		<input checked="" type="checkbox"/>	
Demande de certificat d'utilité		<input type="checkbox"/>	
Demande divisionnaire		<input type="checkbox"/>	
<i>Demande de brevet initiale</i> N° _____ Date ____/____/____ <i>ou demande de certificat d'utilité initiale</i> N° _____ Date ____/____/____			
Transformation d'une demande de brevet européen <i>Demande de brevet initiale</i> N° _____ Date ____/____/____			
3 TITRE DE L'INVENTION (200 caractères ou espaces maximum) Image files constructing tools for cluster configuration.			
4 DÉCLARATION DE PRIORITÉ OU REQUÊTE DU BÉNÉFICE DE LA DATE DE DÉPÔT D'UNE DEMANDE ANTÉRIEURE FRANÇAISE		Pays ou organisation _____ N° _____ Date ____/____/____ Pays ou organisation _____ N° _____ Date ____/____/____ Pays ou organisation _____ N° _____ Date ____/____/____ <input type="checkbox"/> S'il y a d'autres priorités, cochez la case et utilisez l'imprimé «Suite»	
5 DEMANDEUR		<input type="checkbox"/> S'il y a d'autres demandeurs, cochez la case et utilisez l'imprimé «Suite»	
Nom ou dénomination sociale		SUN MICROSYSTEMS, INC	
Prénoms			
Forme juridique			
N° SIREN			
Code APE-NAF			
Adresse	Rue	901 San Antonio Road	
	Code postal et ville	94303	PALO ALTO Californie
Pays		Etats-Unis d'Amérique	
Nationalité		américaine	
N° de téléphone (facultatif)			
N° de télécopie (facultatif)			
Adresse électronique (facultatif)			



BREVET D'INVENTION CERTIFICAT D'UTILITÉ

REQUÊTE EN DÉLIVRANCE 2/2

REMISE DES PIÈCES DATE 30 SEPT 2002 LIEU 75 INPI PARIS N° D'ENREGISTREMENT 0212075 NATIONAL ATTRIBUÉ PAR L'INPI		Réservé à l'INPI	
Vos références pour ce dossier : <i>(facultatif)</i>		SUN 43 (120749)	
6 MANDATAIRE			
Nom		PLAÇAIS	
Prénom		Jean-Yves	
Cabinet ou Société		Cabinet NETTER	
N° de pouvoir permanent et/ou de lien contractuel			
Adresse	Rue	36 avenue Hoche	
	Code postal et ville	75008	PARIS
N° de téléphone <i>(facultatif)</i>		01 58 36 44 22	
N° de télécopie <i>(facultatif)</i>		01 42 25 00 45	
Adresse électronique <i>(facultatif)</i>			
7 INVENTEUR (S)			
Les inventeurs sont les demandeurs		<input type="checkbox"/> Oui <input checked="" type="checkbox"/> Non Dans ce cas fournir une désignation d'inventeur(s) séparée	
8 RAPPORT DE RECHERCHE		Uniquement pour une demande de brevet (y compris division et transformation)	
Établissement immédiat ou établissement différé		<input type="checkbox"/> <input checked="" type="checkbox"/>	
Paiement échelonné de la redevance		Paiement en deux versements, uniquement pour les personnes physiques <input type="checkbox"/> Oui <input type="checkbox"/> Non	
RÉDUCTION DU TAUX DES REDEVANCES		Uniquement pour les personnes physiques <input type="checkbox"/> Requête pour la première fois pour cette invention (<i>joindre un avis de non-imposition</i>) <input type="checkbox"/> Requête antérieurement à ce dépôt (<i>joindre une copie de la décision d'admission pour cette invention ou indiquer sa référence</i>).	
Si vous avez utilisé l'imprimé «Suite», indiquez le nombre de pages jointes			
10 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE (Nom et qualité du signataire) N° Conseil 92-1197 (B) (M) Jean-Yves PLAÇAIS		VISA DE LA PRÉFECTURE OU DE L'INPI M. BLANCANEUX	

La loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés s'applique aux réponses faites à ce formulaire. Elle garantit un droit d'accès et de rectification pour les données vous concernant auprès de l'INPI.

Image files constructing tools for cluster configuration

- 5 The invention relates to a distributed computer system comprising computers or other hardware entities called nodes.

It exists tools to configure a computer individually in order for it to run for example a server similar to a server of another computer. Such tools comprises capture tools and deployment
10 tools. Capture tools enable an administrator to capture in archives the operating system environment, the applications stack and the system configuration from a server. Deployment tools use the content of these archives for the installation of a server (clone) on a single computer.

- 15 However, such tools do not provide efficiency for the configuration of computers in a distributed computer system.

A general aim of the present invention is to provide advances in this matter.

- 20 The invention concerns a software code, adapted to manage a configuration of at least a group of nodes, said software code comprising :

- input code, capable of receiving a set of model configuration files, adapted to create and configure at least partially a data model, said data model defining hardware entities and logical entities for a group of nodes,
- 25 - generator code, capable of generating first node data in cooperation with the data model as at least partially configured,
- install code, capable of installing a specific environment in a machine having at least partially the configuration of the nodes of the group, in order to create an archive object using first node data,
- 30 - configuration code, enabling a user to complete the configuration of the data model dynamically and generate second node data, and
- further install code, capable of installing a specific environment in a group of nodes, in order to create a deployable object from the archive object and the second node data, and in order to configure nodes of the group of nodes in deploying said deployable object.



The invention also concerns a method of managing a configuration of at least a group of nodes comprising the following steps :

- a- receiving as input a set of model configuration files and configuring at least partially a data model, said data model defining some hardware entities and some logical entities for a group of nodes,
- b- generating first node data in cooperation with the data model at least partially configured,
- c- installing a specific environment in a machine having at least partially the configuration of the nodes of the group in order to create an archive object using first node data,
- d- completing the configuration of the data model dynamically and generating second node data,
- e- installing a specific environment in a group of nodes in order to create a deployable object from the archive object and the second node data, and in order to configure nodes of the nodes group in deploying said deployable object.

The invention also concerns a software product for performing the hereinabove method.

Other alternative features and advantages of the invention will appear in the detailed description below and in the appended drawings, in which :

- figure 1 represents a distributed computer system comprising two nodes;
- figure 2 represents a group of nodes forming a cluster;
- figure 3 is a partial illustration of the machines having the software management and configuration tools used for the invention;
- figure 4 is a complete illustration of a build environment and of a deployment environment according to the invention;
- figure 5 is a functional diagram representing the input of the software management and configuration tools in the invention;

- figure 6 is a functional diagram representing the output of the software management and configuration tools in the invention;

5 - figure 7 represents a flow chart of a generic flash archive creation process according to the invention;

- figure 8 represents a flow chart of a configured flash archive creation process according to the invention;

10 - figure 9 represents a flow chart of a flash archive deployment process according to the invention,

-figure 10 represents a data model used by the software management and configuration tools.

15 The drawings form an integral part of the description embodiments of the present invention.

This invention also encompasses embodiments in software code, especially when made available on any appropriate computer-readable medium. The expression "computer-readable medium" includes a storage medium such as magnetic or optic, as well as a
20 transmission medium such as a digital or analog signal.

This invention may be implemented in a network comprising computer systems. The hardware of such computer systems is for example as shown in Fig. 1, where in the computer system 10:

25 - 1-10 is a processor, e.g. an Ultra-Sparc (SPARC is a Trademark of SPARC International Inc);

- 2-10 is a program memory, e.g. an EPROM for BIOS;

- 3-10 is a working memory for software, data and the like, e.g. a RAM of any suitable technology (SDRAM for example); and

30 - 7-10 is a network interface device connected to a communication medium 8, itself in communication with other computers such as computer system 11. Network interface device 7-10 may be an Ethernet device, a serial line device, or an ATM device, inter alia. Medium 8 may be based on wire cables, fiber optics, or radio-communications, for example.

The computer system 10 may be a node amongst a group of nodes in a distributed computer system. The other node 11 shown in figure 1 comprises the same components as node 10, the components being designated with the suffix 11. The node 11 further comprises a mass memory 4-11, e.g. one or more hard disks. The nodes having hard disk are designated as diskfull and other nodes having no hard disk are designated as diskless.

Thus, node 10 is considered as a diskless node and node 11 is considered as a diskfull node.

Data may be exchanged between the components of Figure 1 through a bus system 9-10, respectively 9-11, schematically shown as a single bus for simplification of the drawing. As is known, bus systems may often include a processor bus, e.g. of the PCI type, connected via appropriate bridges to e.g. an ISA bus and/or an SCSI bus.

Figure 1 depicts two connected nodes.

Figure 2 shows an example of groups of nodes arranged as a cluster C. The cluster may have a group of master eligible nodes and optionally at least a group of non master eligible nodes. By way of example only, the cluster comprises a group of master eligible nodes having a master node NM adapted to manage the nodes of the cluster and, for reliability reasons, a vice-master node NV, and a group of non master eligible nodes comprising the other nodes N2, N3 ... Nn-1 and Nn. Each node is connected to the network 31 which may be redundant. The network 31 is adapted to link nodes of a cluster between them. These interconnections between nodes of a cluster may be done using a switch. The qualification as master or as vice-master should be viewed as dynamic: one of the nodes acts as the master (resp. Vice-master) at a given time. However, for being eligible as a master or vice-master, a node needs to have the required "master" functionalities. A node being diskfull is considered to have at least partially these master functionalities. When informed about the master node failure, the vice-master node may replace the failed master node and becomes the master node.

Several definitions may be useful to understand the following description :

As seen, a *node* is a logical entity which may involve:

- . a main cPCI board (e.g. CP2060, CP1500, ...) or a Netra machine (e.g. Netra 20)
- . Ethernet NICs

- . optionally disk(s)
- . optionally PMC (PCI Mezzanine Cards, as e.g. co-processor)
- . optionally Input /Output front and/or rear cards.

In a cluster, a node may be elected as the master node of the cluster and thus managing the other nodes. In a cluster, there may be one or more master eligible node (s) in the group of master eligible nodes.

A *diskless node* is a node without storage capabilities. Diskless nodes are always non master eligible nodes. This kind of node runs/uses an Operating System, data and software exported by the master node of the cluster.

A *diskfull node* is a node which has storage capabilities. Diskfull nodes can be Master Eligible or Non Master Eligible.

A *dataless node* is a Non Master Eligible diskfull node. This kind of node shares data and software exported by the master node of the cluster. Contrary to a diskless node, this kind of node uses an Operating system stored on a local disk, boots locally and has local storage capabilities.

A *nodes group* defines a set of nodes which share the role (Master Eligible or Non Master Eligible), the Operating System, the Architecture (e.g. SPARC), the class, the software, the foundation services configuration and the peripherals capabilities (e.g. same disk capabilities, same i/o cards capabilities). Furthermore every nodes of a Master Eligible node group may have the same disk(s) layout(s), the same file systems definitions. The master eligible nodes group (one per cluster) embeds diskless nodes groups environments, if some are defined.

Flash archives are a set of data created with Flash tools which are further described in the following documentation : Solaris 9 Installation Guide. (ref 806-5205-10) and comprising as known for example the system image of each node of a cluster, also called cluster image. This system image is a sort of image file (called archive file) which may comprise operating system files, cluster configuration files, user application software and data. When installed on a target machine (e.g. in epROM or flash memory), the image file is capable (after

suitable initialization, like re-booting) to have it operate consistently in the desired hardware and software configuration. This system image may be replicated (or deployed) on multiple cluster. The flash archive is deployed using replication tools such as the Jumpstart (trade mark) scripts of Sun microsystem for example further described in the hereinabove cited
5 documentation : Solaris 9 Installation Guide.

There may be three types of flash archives :

- *Generic Flash archive*: this archive cannot be deployed. It is a cluster-independent flash archive. It can be used as basis to generate deployable Flash archive for a set of clusters
10 sharing the topology defined in the cluster image.
- *Configured Flash archive*: generic Flash archives which embeds user application data and scripts to be installed on the cluster nodes at deployment time.
- *Deployable Flash archive*: generic or configured Flash archive adapted to a specific cluster (or a specific site).

15

A *Software Load (SWL)* is a collection of deployable objects (i.e. Flash archives dedicated for a specific cluster).

The *Software Management and Configuration Tools* according to the invention is a Sun
20 product comprising software tools adapted to configure, build and deploy the software Load (SWL) on a cluster.

Figure 3 represent a single machine S on which is installed and is running the Software Management and Configuration Tools of the invention. As illustrated, machine S is linked
25 to a machine which represents either a machine C* called a master system at flash archive creation time (more precisely at software load creation time), either a group of running nodes (e.g. a cluster) at flash archive deployment time (more precisely at software load deployment time). The master system (also called a prototype machine) is a machine equivalent to a master eligible node(s) or to a dataless node(s) and from which the Flash archives (and the
30 software load) are built. The master system is a machine comprising the same architecture (e.g. SPARC) and the same disk configuration than the nodes of the group of nodes in the cluster on which the flash archive (more precisely the software load) is to be deployed.

Figure 4 represents another embodiment of installation of the Software Management and Configuration Tools. The Software Management and Configuration tools are installed partially on a build server S1 and on an install server S2 linked to the master system CMS for flash archive creation and on several install server S3, S4 for example linked to cluster
 5 C1, C2, C3, C4 for flash archive deployment. S1, S2, CMS define the flash archive creation environment (FAC) and each install server S3, S4 defines a flash archive deployment site.

A build server may be a machine, e.g. a standard Solaris server, comprising tools to manage software data to create the flash archive and then the software load, e.g. in preparing these
 10 software data. User application data and scripts may also be prepared by the build server in order to be added to flash archives. This build server may work in relation with an install server. An install server is a machine used as host for the environments and comprising tools to install a specific environment on a master system and on a cluster nodes. Such a server is used, after such specific environment installation and gathering of software data by the
 15 build server, to create the generic flash archives and the software load or to deploy them on the nodes of a nodes group. Configured flash archives may also be created by the install server if user configuration data and scripts are added to the generic flash archives. The software loads may all be in a central repository.

20 Figure 5 describes the inputs used by the Software Management and Configuration Tools. Other inputs may be used according to the invention. References to Solaris and Sun products are only examples.

The operating system standard Sun Microsystems tools 100 comprise the Jumpstart tools
 25 and the flash tools. The Software Management and Configuration Tools use Jumpstart tools for the installation of a specific environment on

- the master systems using software data for the flash archive creation and on
- the cluster nodes for the deployment and the installation of the flash archives.

The Software Management and Configuration Tools use the Flash tools to create the Flash
 30 archives.

The Software code may comprise different software 130 from different products and user applications, e.g. Sun products. These software may comprise several packages and patches,

e.g. from Solaris add-on products. For example, some packages and patches come from the Solaris distribution used by Jumpstart.

5 The configuration data 110 used by the software management and configuration tools may comprise cluster data, machine data, network data, jumpstart data and nodes group data such as software data and configuration data.

The User configuration data 120 comprise application configuration data defined by any user and the scripts able to install them on the cluster.

10

All these inputs 100, 110, 120 and 130 are organized according to a classes diagram created and configured as described in reference to figure 10.

15 Such a classes diagram (called a data model) is adapted to define a cluster model 800, a machine model 700 and a network model 900. Links between classes in figure 10 represent the relationships between classes : thick lines represents containments, fine lines represent references, dotted lines represent inheritances.

20 The machine model 700 describes all the hardware elements of the cluster and their connections. The cluster topology is thus defined. In this model, the machine is seen as a hierarchical collection of hardware elements. The outer part of the machine is the shelf (or chassis), the inner part of the machine are the node peripherals (disk, ethernet interface NIC, ...). The model allows the definitions of several shelves per cluster. Thus, the shelf class 705 contains a switch class 706 and a drawer class 704. The drawer class 704 contains a disk
25 class 703 to define the disk(s) used and to a board class 708 to indicate which board is used. The disk class 703 contains the slice class 702 itself containing the file system class 701 to define the file system of a disk slice. The board class contains the ethernet interface class (nic class) 709 to define possible multiple ethernet interfaces for a board. To define topology of a cluster, the ethernet interface class 709 is linked to the switch class 706 indicating that
30 a ethernet interface is linked to a defined switch and more precisely to a defined port of a switch. Indeed, the switch class 706 contains the port class 707. A reference link exists between the port class and the ethernet interface class.

A cluster model 800 describes the cluster nodes and nodes group. It is the logical view of the cluster. It comprises different classes hereinafter described.

- 5 The cluster class 807 contains the nodes group class 801. This nodes group class 801 itself contains the nodes class 802. The nodes class inherits an operating system class 810, which may contains e.g. Solaris or any other operating systems for a node. Each node group is referred to a software class 803 which defines the software installed for each node of a nodes group. This software class 803 inherits the correction class 811 defining e.g. a patch enabling modification of files in a software distribution for a node group and the installation
- 10 unit class 812 defining e.g. the package for Solaris environment being a software distribution. User defined software distribution, patches and packages may be defined optionally in these classes. Software distribution, patches and packages may also be defined for other environment than Solaris in these classes.
- 15 Each nodes group of the nodes group class 801 is also referred to a configuration class (not shown) itself referred to a file class. The configuration class defines user configuration data associated to a nodes group. The file class defines any data file or install scripts of user configuration data associated to a nodes group.
- 20 Moreover the software class 803 contains a software repository class 805. This class 805 may be used to store the software to be integrated in a flash archive and to define which software of the software repository is adapted to produce a given flash archive for a given nodes group. For the configuration class not shown, a softwareload repository class (not shown) may comprise the flash archives themselves once created. This last class and class
- 25 805 provides a tool for reproducibility of flash archives as described hereinafter.

The node group class 801 is referred to the service class 804. This class 804 is adapted to provide pre-configured services, e.g. for the management of a cluster or for the constitution of a cluster (Cluster monitor membership).

30

The cluster class 807 further contains the domain class 806. Each cluster comprises a domain number.

The network model 900 is adapted to configure the IP addressing schema of the network. The network model 900 comprises different classes defining the network. These different classes are linked on one hand to the domain class 806 and on the other hand to the switch class 706 and the ethernet interface class 709. These different classes comprise the IP address class 903, the network class 902 and the router class 901.

Each of the three models 700, 800, 900 is configured by a configuration file, e.g. a cluster configuration file for the model 800, a machine configuration file for the model 700 and a network configuration file for the model 900. Each configuration file may be fully defined or partially defined according to the process of figures 7, 8 or 9 used as hereinafter seen.

Some input data may be added or specified in an incremental way, i.e. configuration files for these models may be added as input data at flash archive creation or at deployment process as described in figures 7, 8, 9. In other words, some input data may be added dynamically according to the process used.

Figure 7 represents the generic flash archive creation process according to the invention and managed by the software management and configuration tools.

The process of flash archive creation is based on a set of configuration files taken as input by the software management and configuration tools. These configuration files comprise

- a cluster configuration files comprising a cluster definition defining the cluster nodes, the cluster domain, the nodes group and the foundation services associated to each nodes group; this file configures the cluster model of figure 10,
- a machine configuration file defining the machine (or computers) running the cluster; this file configures partially the machine model of figure 10 (e.g. with file system definitions).

Further configuration files are used ;

- nodes group configuration files comprising software configuration files for foundation services (i.e. packages and patches) which rely on predefined configuration files and, optionally, for user defined software,
- jumpstart configuration file defining a jumpstart profile for each diskfull node. It defines the operating system standard tools (or packages for Solaris) to install on such nodes.

In a first operation 402, the build server is adapted, from the input configuration files and for each nodes group, to gather the software, to generate software install scripts, to generate Jumpstart profile for master eligible and dataless nodes groups only. Operation 402 also generate a software load descriptor. Once operation 402 has succeeded, operation 404 takes
 5 the node data generated by operation 402 and copy them into a directory.

At operation 406, the install server is adapted to generate a Jumpstart installation environment for a master system from the data in the directory generated by operation 404. This operation is run once per group, each nodes group is processed individually; in other
 10 words, each master system has the software data of the nodes group targeted. Once the Jumpstart installation is finished, the master system reboots and the user may perform any customization at operation 408. At operation 410, the install server generates a generic Flash archive from the installed master system. The generic Flash archive created may comprise two sections:

- 15 - one binary section which corresponds to the system image, i.e. being the 'archive files' section of the Flash archive;
- one ascii section which holds the signature of the software load descriptor generated at operation 402. This signature enables to identify the flash archives in order to perform consistency checks in the next steps of figure 8 and 9, to describe the content of the flash
 20 archive and to interrogate and/or display the content of the flash archives.

The software load may be created at this time.

The generic flash archive creation ends. Though the operations are processed, as indicated in figure 7, in the different servers (build server and install server), they may be processed
 25 in a single server having the software management and configuration tools.

Figure 8 illustrates an optional process of the invention : the configured flash archive creation process. The Flash archive configuration consists in adding or replacing configuration data sections to an existing Flash archive. This configured flash archive
 30 creation is based on the following inputs taken by the software management and configuration tools:

- a flash archive (generic according to the process of figure 7 or already configured according to the process of figure 8) stored in the central repository,

- user data given as a set of files called user defined configuration data, these data being added in the configuration class of the cluster model,
- installation scripts for these user data, these scripts being added in the file class of the cluster model.

5

The control of the order of scripts execution and the configuration of data and scripts are allowed by user defined configuration files.

At operation 502, the build server is adapted to gather user defined configuration data, e.g. a set of files, and user defined configuration data installation scripts given as parameters. These scripts and data are configured at deployment time as seen in figure 9. These scripts are aimed at installing the user data at deployment time. In an embodiment, all the nodes groups are processed in one operation. Once operation 504 has succeeded, the data generated by this operation 504 are copied into a directory. The data generated by operation 402 are also exported as they are in the directory. Operation 506 generates ascii sections from the User defined data. A single section may include all the nodes group configuration data and scripts. If a Flash archive is given as input, e.g. a generic flash archive from step 410 figure 7 or a already configured flash archive, the configuration sections are replaced by the new ones. This operation 506 runs once per nodes group, i.e. each nodes group is processed individually.

20

Configured flash archives are thus created at operation 508. The creation process ends. The configured flash archives may be stored in the central repository for reproducibility of deployment process of figure 9.

25

All these data inputs of operation 502 are optional. If none of these data inputs exist, no new sections are created or replaced. The configuration data and scripts are configured at the nodes group level so that the associated flash archives can be configured.

30

Figure 9 represents the flash archive deployment process according to the invention and managed by the software management and configuration tools. The install server and the build server are linked to the nodes group on which the flash archives are to be deployed. Some data are given in input to perform this process comprise:

- the flash archives coming from the processes of figure 7 or 8,
and the configuration data :

- the configuration files, configuring the model of figure 10, being:

- 5 a cluster configuration file being logical cluster definition which defines cluster nodes, nodes groups and foundation service list; the foundation services comprise the different services that a node in a cluster may need (Reliable Boot Service (RBS), Cluster Membership Monitor (CMM) ...),
- a machine configuration file which defines the machine running the cluster,
- 10 a network configuration file which defines the networks parameters needed by the cluster.

Different consistency checks are accomplished in regard with the foundation services to configure. Configuration data may be imported from the flash archive configuration process in input (more specifically from the software repository class) : for a configured flash archive, these configuration data may indicate when to run the user install scripts for
15 applications as described hereinafter.

At operation 602, the build server generates the foundation services configuration files and the operating system configuration files according to the input being e.g. cluster, machine and network configuration files of the data model. The foundation services configuration
20 files are, for example, network configuration files or DHCP configuration files. Operation 602 may be run once per software load, i.e. all the nodes groups are processed in one operation. A consistency check may be done with imported configuration data. Once operation 602 has succeeded, operation 604 exports data from operation 602 and copy them into a directory.

25

At operation 606, the install server adds sections to the Flash archive given as input. These sections include the exported data of operation 604 copied in the directory. The software load descriptor stored in the software load section as seen in figure 7 is replaced in order to take into account the complete cluster definition. The flash archives are now deployable
30 flash archives. Another consistency check may be done using the archive and the software load data. At operation 608, the install server generates a Jumpstart environment for each node of the nodes group given as parameter, i.e. each nodes group may be processed



individually. The install server deploys then the flash archives, e.g. in the form of a software load, in each node of group of nodes.

At operation 610, the nodes of the cluster are configured using the deployed flash archives.

5 Then, if the flash archive defines no user configuration data and scripts, the process continues at operation 612 for a reboot of the nodes of the nodes group, the optional operations 614, 616 and 618 (surrounded with a dotted line) are not processed and then it ends with the run of the cluster. If the flash archive defines user configuration data and scripts, the optional operations may be executed according to the following different cases:

10

- a second reboot may have been configured and in this case, the foundation services (also called cluster services) are not started yet. The user configuration data are configured at operation 614 and the second reboot starts the foundation services at operation 616. The process may then configure other user configuration data at operation 618 and ends with the run of the cluster or directly ends with the run of the cluster.

15

- the reboot of operation 612 has started the foundation services and the user configuration data are configured at operation 618. The process may then ends with the run of the cluster.

20 Configuration data in the configuration class of figure 10 are used to indicate at which operation running the user install scripts.

These double reboot operations is used to configure any kind of nodes group, e.g. master eligible group of nodes or dataless group of nodes of a cluster, using user configuration data.

25 To take into account cluster dependent data, process of figure 8 may be run again to change the content of the configuration data sections in the flash archives.

30 The first tools used to prepare the final configuration data may be in the build server or in the install server connected to the cluster. The second tools incorporating the configuration data into the Flash archives and creating the cluster Jumpstart deployment environments may be in the install server. The final configuration data can then be merged into a centralized software load repository. To communicate the software load data between the different environments, import and export operations are used.

Figure 6 recapitulates the different outputs provided by the processes of figures 7, 8, 9.

The flash archives are the main outputs of the software management and configuration tools. The flash archive 210 may be a self-describing deployment object. At least the three type of
 5 flash archives per software load are managed by the software management and configuration tools. Repositories 200 such as software repository or software load repository are other outputs and enable to rebuild an identical or similar flash archive with stored data if needed (e.g. if a flash archive is destroyed). Other outputs comprise the master systems jumpstart environment and the cluster nodes jumpstart environment or other environment adapted to
 10 provide tools 230 to deploy flash archive. Outputs also comprise cluster configuration files 220 as foundation services configuration files and operating system configuration files.

The invention enable flash archives to be reproducible and configured depending on cluster configuration and/or user defined configuration data. Flash archives are thus generated and
 15 deployed using single software tools.

Using the different models (cluster, machine and network models) having repositories, it is possible to execute independently each process of figures 7, 8, 9 by providing consistency checks. The invention further enables a simplification and a standardization of flash archive
 20 creation and deployment process for a nodes cluster.

The invention is not limited to the hereinabove features. Other environments may be used.

Some operations may also be used for supplementary configurations. Thus, the double
 25 reboot operations of figure 9 may also be used to manage internally the configuration of some services. For example, for a given service if the configuration is done during the cluster configuration and if this service requires e.g. another services missing, the double reboot operations enables to configure the given service on the nodes of the node group.

30 This invention also covers the software code for performing the method of figure 7, 8 and 9, especially when made available on any appropriate computer-readable medium. The expression "computer-readable medium" includes a storage medium such as magnetic or optic, as well as a transmission medium such as a digital or analog signal. The software code



basically includes separately or together, the codes defining the build server, the install server and the master system.

α

Claims

1. A software code, adapted to manage a configuration of at least a group of nodes, said software code comprising :
 - 5 - input code, capable of receiving a set of model configuration files, adapted to create and configure at least partially a data model, said data model defining hardware entities and logical entities for a group of nodes,
 - generator code, capable of generating first node data in cooperation with the data model as at least partially configured,
 - 10 - install code, capable of installing a specific environment in a machine having at least partially the configuration of the nodes of the group, in order to create an archive object using first node data,
 - configuration code, enabling a user to complete the configuration of the data model dynamically and generate second node data, and
 - 15 - further install code, capable of installing a specific environment in a group of nodes, in order to create a deployable object from the archive object and the second node data, and in order to configure nodes of the group of nodes in deploying said deployable object.
2. The software code of claim 1, wherein the first node data and/or second node data are
20 stored in a repository enabling the re-creation of an archive object and/or a deployable object, as required.
3. The software code of claim 2, wherein the first node data comprise software data and a software install script for a group of nodes.
- 25 4. The software code of claim 2, wherein the second node data comprise operating system files and service configuration files for a group of nodes.
5. The software code of claim 4, wherein the install code is further adapted to create a
30 configured archive object in adding user defined configuration data and user install scripts to an archive object.



6. The software code of claim 5, wherein the install code is further adapted to create a new configured archive object in adding new user defined configuration data and new user install scripts to an already configured archive object.
- 5 7. The software code of claim 5, wherein the further install code is further adapted to configure nodes of the group of nodes, to reboot the nodes, to configure the user defined configuration data and to run the nodes.
8. The software code of claim 5, wherein the further install code is further adapted to
10 configure nodes of the group of nodes, to process to a first reboot the nodes, to configure the user defined configuration data, to process to a second reboot and to run the nodes.
9. The software code of claim 1, wherein the input code is adapted to receive the set of
15 model configuration files, as comprising a logical configuration file for nodes groups and a hardware configuration file, and is further adapted to receive a network configuration file in order to complete the data model.
10. The software code of any of the preceding claims, wherein the data model as at least
20 partially configured comprises a hardware model linked to a logical model for groups of nodes.
11. The software code of any of the preceding claims, wherein the completed data model
25 comprises a hardware model linked to a logical model for groups of nodes and to a network model.
12. The software code of any of the preceding claims, wherein the data model is organized
in classes.
13. The software code of any of the preceding claims, wherein the machine is a prototype
30 machine having at least partially the configuration of nodes of a group of node for which an archive object is created.

14. The software code of any of the preceding claims, wherein the deployable object is a deployable flash archive.

5 15. The software code of any of the preceding claims, wherein the archive object is a flash archive.

16. A method of managing a configuration of at least a group of nodes comprising the following steps :

- 10 a- receiving as input a set of model configuration files and configuring at least partially a data model, said data model defining some hardware entities and some logical entities for a group of nodes,
- b- generating first node data in cooperation with the data model at least partially configured,
- 15 c- installing a specific environment in a machine having at least partially the configuration of the nodes of the group in order to create an archive object using first node data,
- d- completing the configuration of the data model dynamically and generating second node data,
- 20 e- installing a specific environment in a group of nodes in order to create a deployable object from the archive object and the second node data, and in order to configure nodes of the nodes group in deploying said deployable object.

17. The method of claim 16, wherein the first node data of step b- and/or second node data d- are stored in a repository enabling the recreation of an archive object in step c-and/or a
25 deployable object in step e-, as required.

18. The method of claim 17, wherein the first node data of step b- comprise software data and a software install script for a group of nodes.

30 19. The method of claim 17, wherein the second node data of step d- comprise operating system files and service configuration files for a group of nodes.



20. The method of claim 16, wherein step c- further comprises creating a configured archive object in adding user defined configuration data and user install scripts to an archive object, said configured archive object being used in step e- to create a deployable object.

5 21. The method of claim 20, wherein step c- further comprises creating a new configured archive object in adding new user defined configuration data and new user install scripts to an already configured archive object.

10 22. The method of claim 20, wherein step e- further comprises configuring nodes of the group of nodes, to reboot the nodes, to configure the user defined configuration data and to run the nodes.

15 23. The method of claim 20, wherein step e- further comprises configuring nodes of the group of nodes, processing to a first reboot the nodes, configuring the user defined configuration data, and processing to a second reboot and to run the nodes.

20 24. The method of claim 16, wherein step a- comprises receiving the set of model configuration files, as comprising a logical configuration file for nodes groups and a hardware configuration file, and further comprises receiving a network configuration file in order to complete the data model.

25 25. The method of any one of claims 16 to 24, wherein the data model as at least partially configured of step b- comprises a hardware model linked to a logical model for groups of nodes.

26. The method of any one of claims 16 to 25, wherein the completed data model of step d- comprises a hardware model linked to a logical model for groups of nodes and to a network model.

30 27. The method of any one of the claims 16 to 26, wherein the data model is organized in classes.

28. The method of any one of claims 16 to 27, wherein the machine of step d- is a prototype machine having at least partially the configuration of nodes of a group of node for which an archive object is created.

5 29. The method of any one of claims 16 to 28, wherein the deployable object of step e- is a deployable flash archive.

30. The method of any one of claims 16 to 29, wherein the archive object of step c- is a flash archive.

10

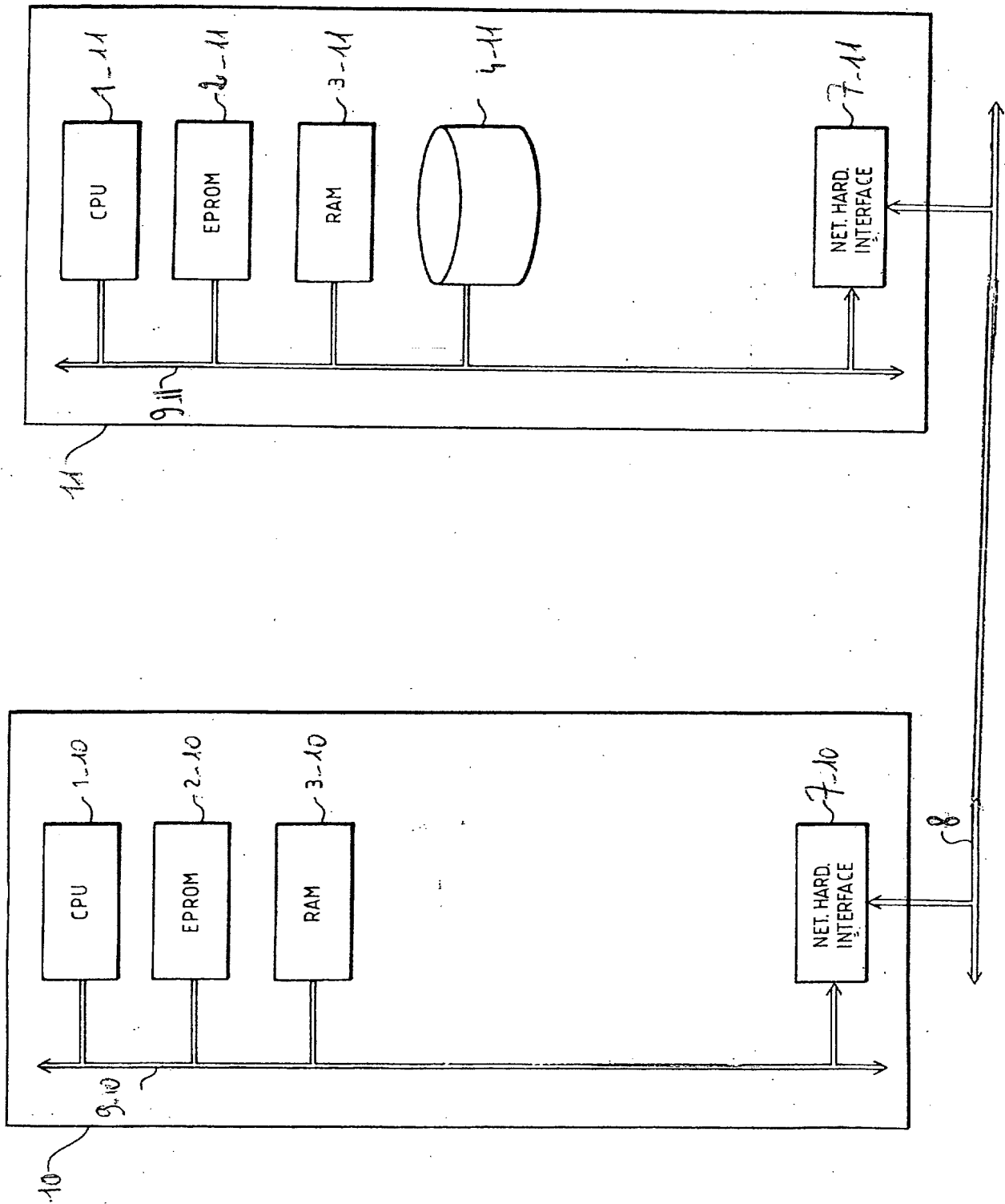
31. A software product for performing the method as claimed in any of claims 16 to 30.

α (21 pages)

CABINET NETTER

118

FIG. 1



2/8

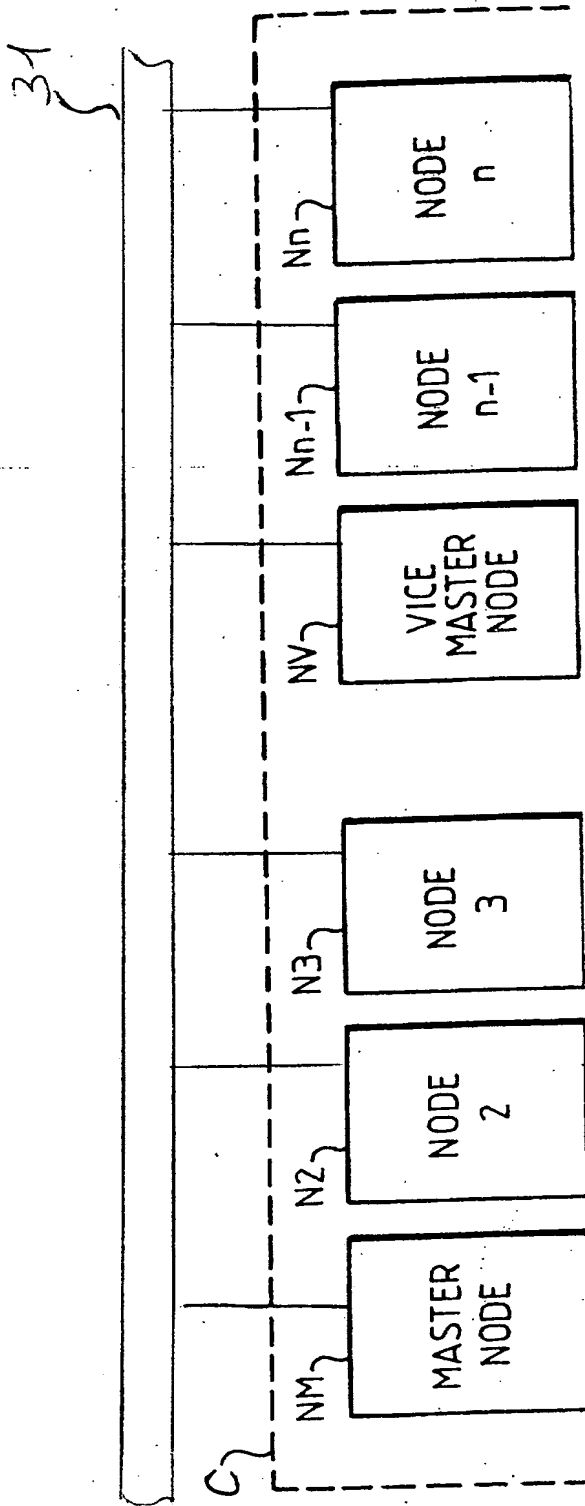


FIG 2

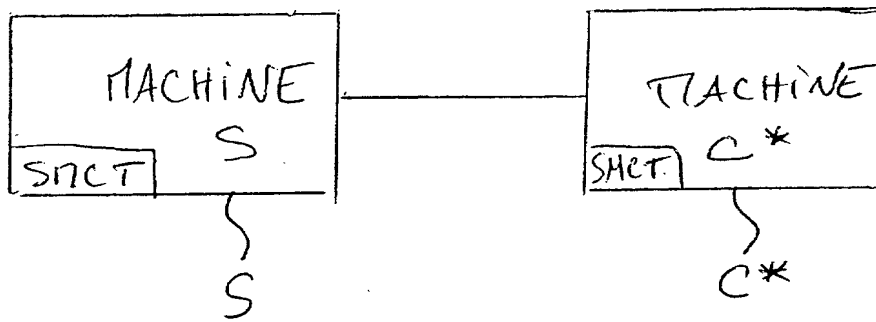


FIGURE 3

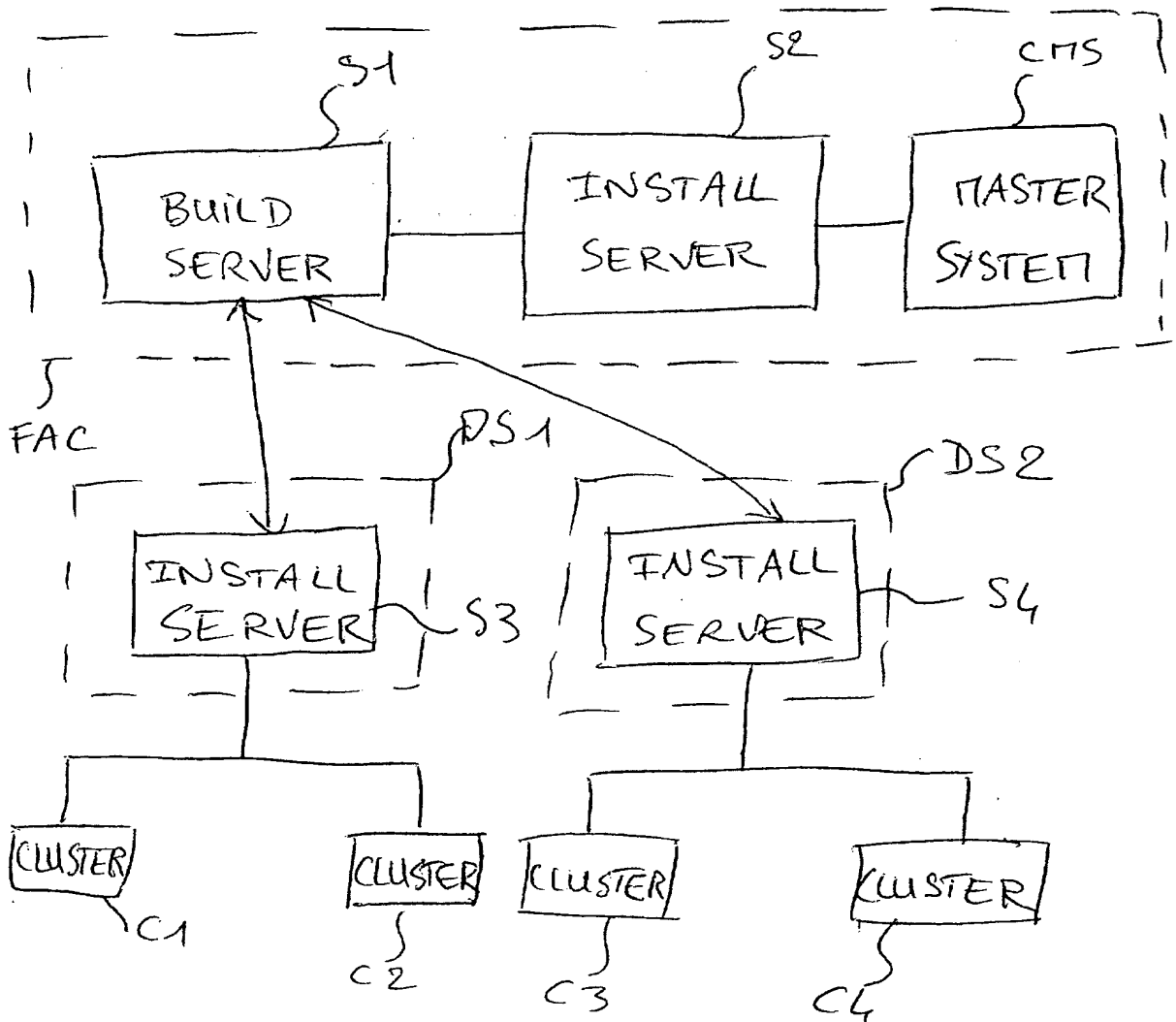


FIGURE 4

4/8

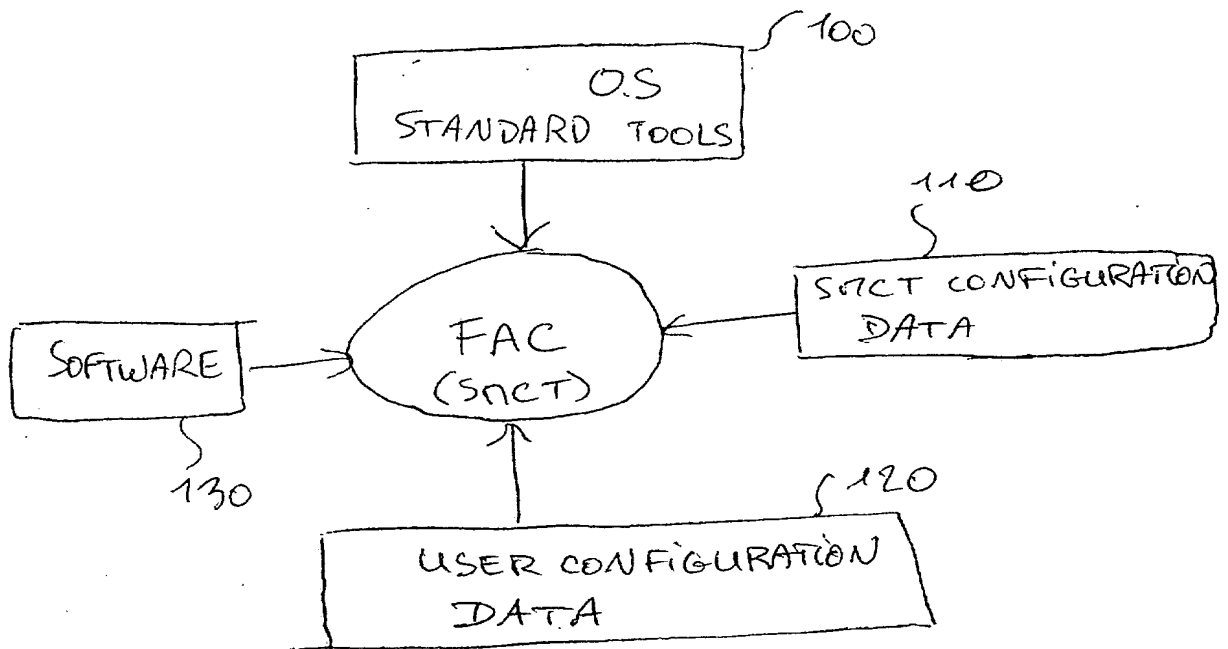


FIGURE 5

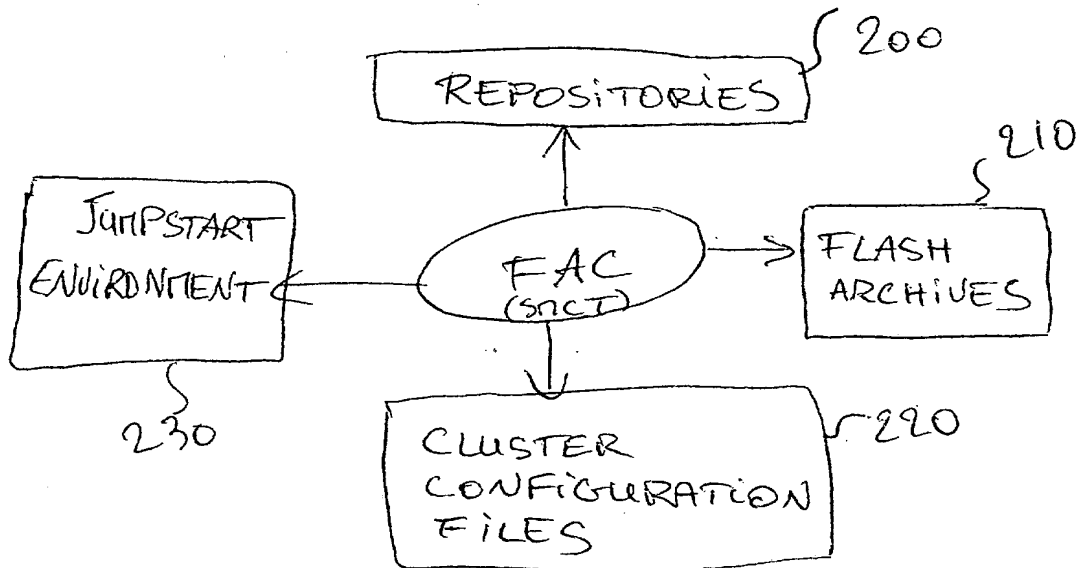


FIGURE 6

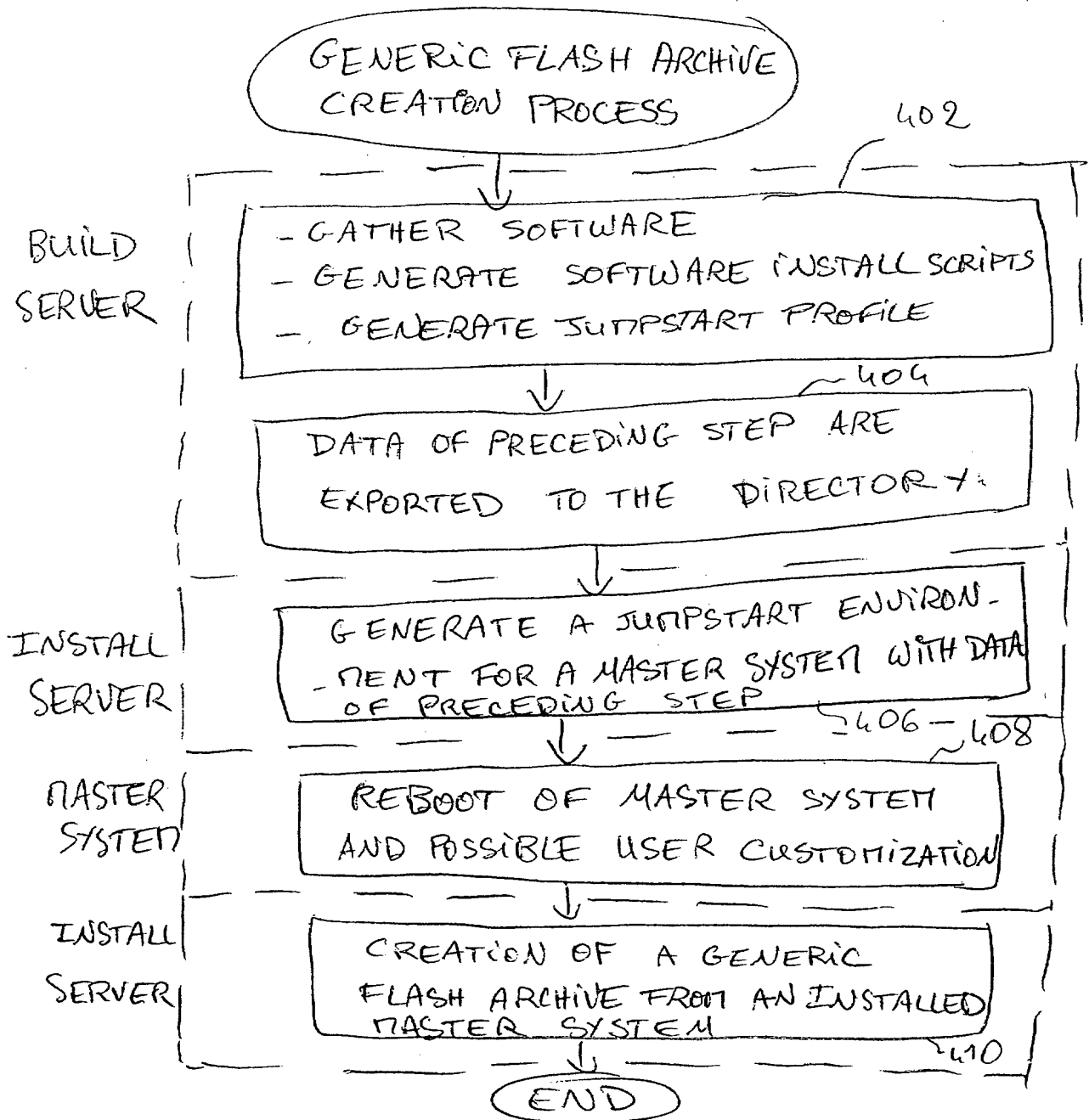


FIGURE 7

6/8

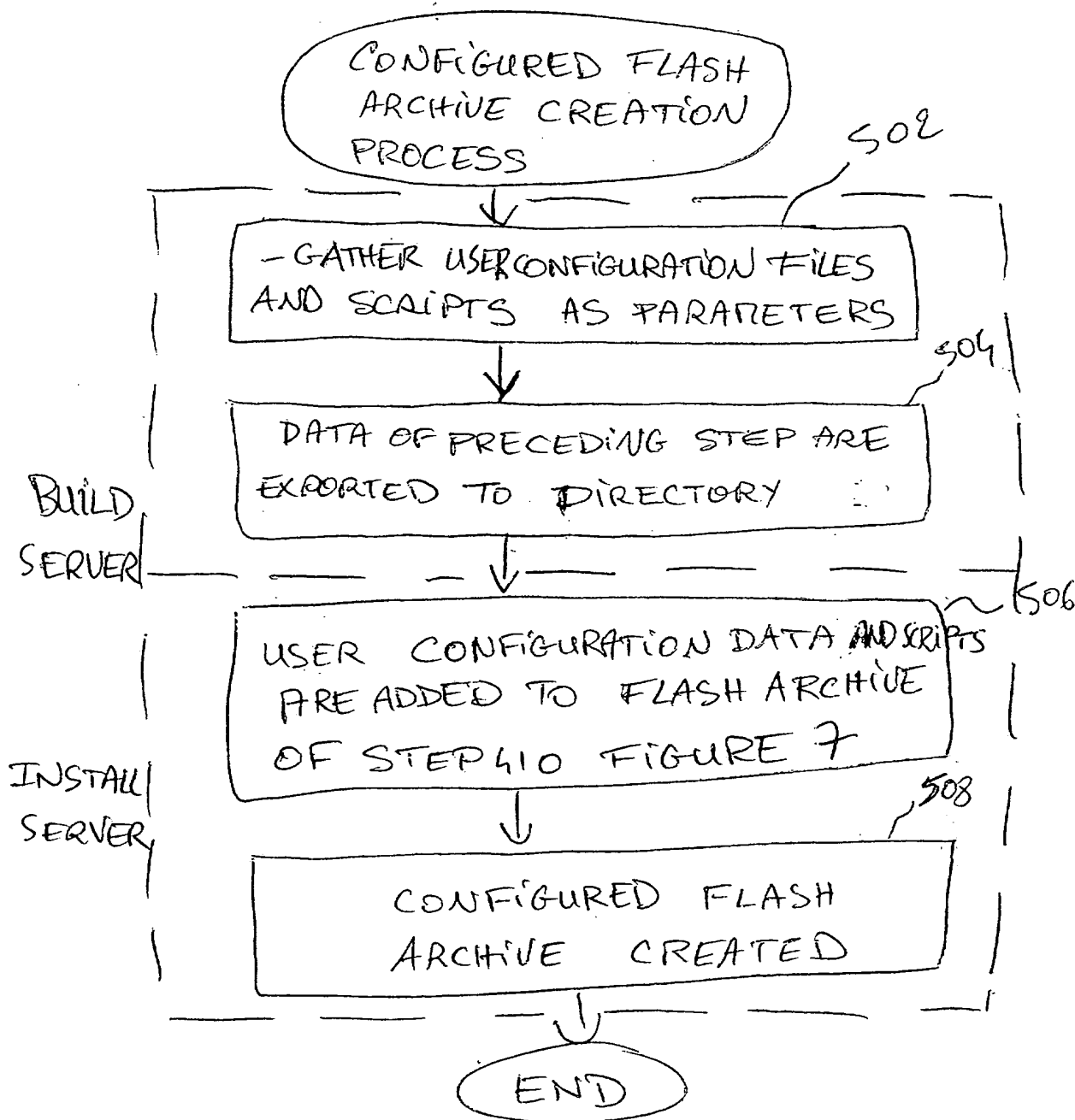
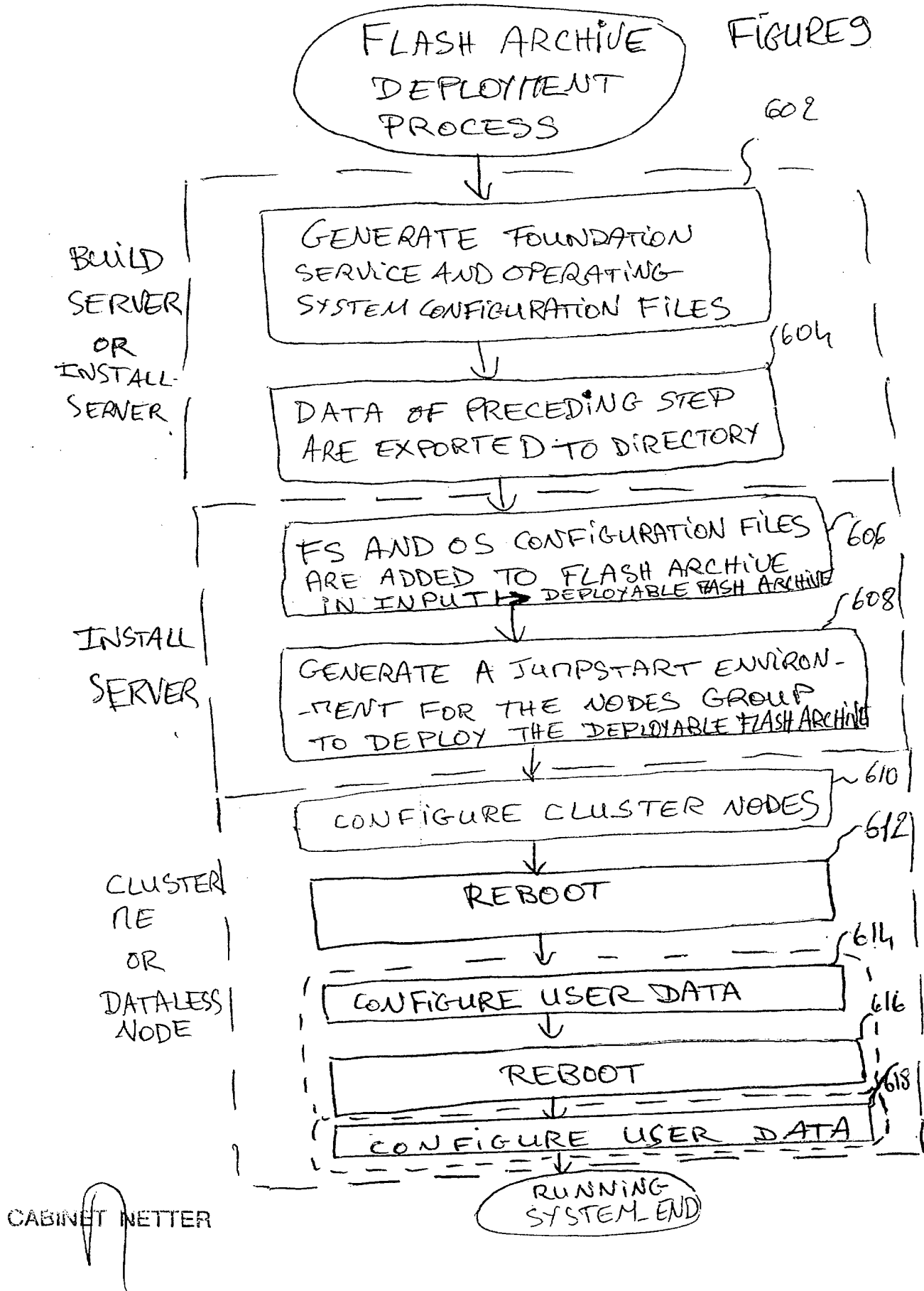


FIGURE 8

4/8



8/8

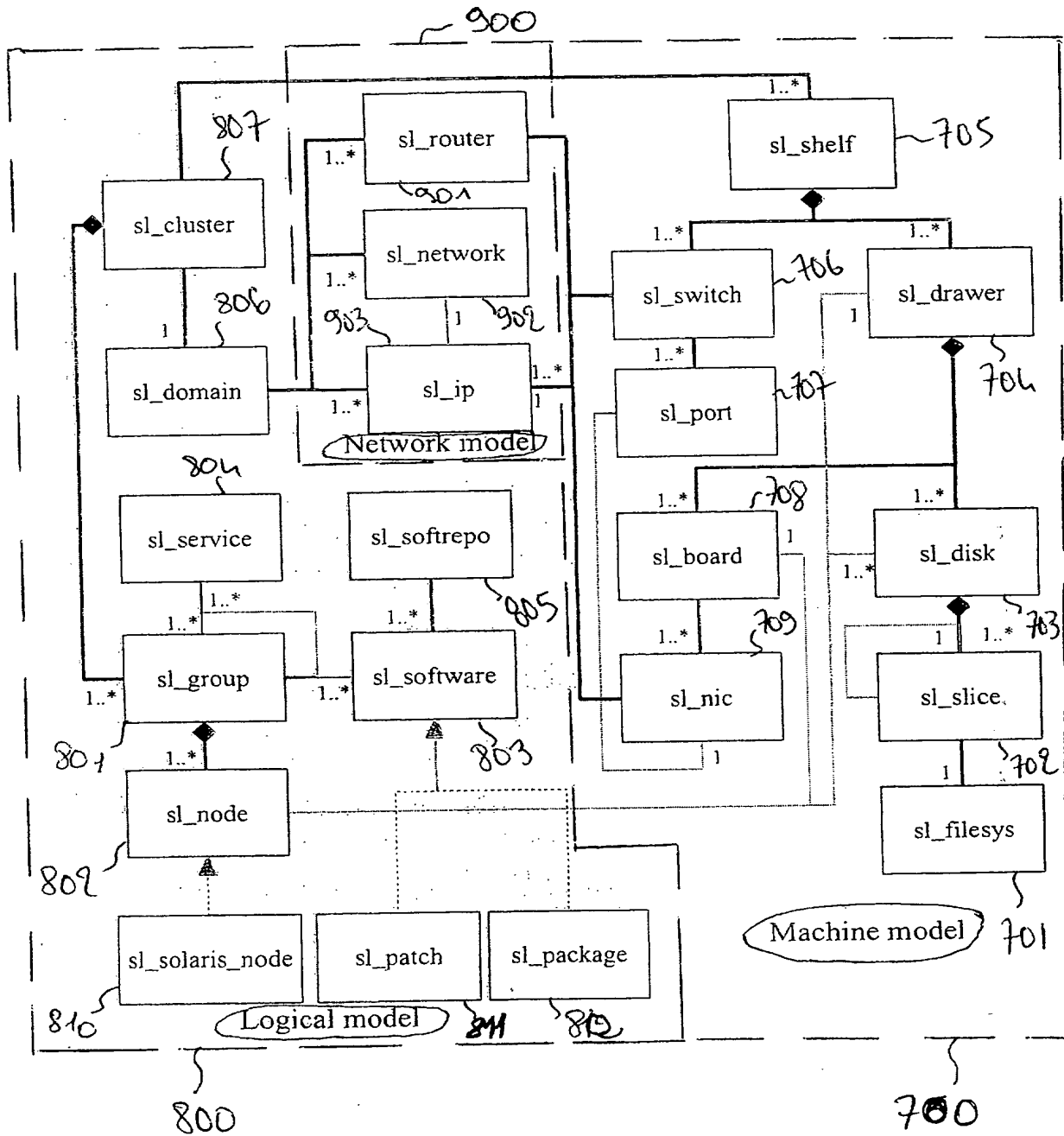


FIG 10

THIS PAGE BLANK (USPTO)